Online, Standards-Based, Formative Assessment Conference Proceedings

Mary Axelson

December 2005

Appalachia Educational Laboratory
at

EDVANTIA™
Partners in education. Focused on results.

Edvantia was founded in 1966 as the Appalachia Educational Laboratory, Inc. (AEL); on September 1, 2005, AEL became Edvantia, Inc. The Regional Educational Laboratory for the Appalachian region is known as the Appalachia Educational Laboratory at Edvantia.

Edvantia is a nonprofit education research and development corporation, founded in 1966, that partners with practitioners, education agencies, publishers, and service providers to improve learning and advance student success. Edvantia provides clients with a range of services, including research, evaluation, professional development, and consulting.

For information about Edvantia research, products, or services, contact

EDVANTIA™

P.O. Box 1348, Charleston, WV 25325 • 304.347.0400 • 800.624.9120 • fax 304.347.0487
One Vantage Way, Suite D-210, Nashville, TN 37228 • 615.565.0101 • fax 615.565.0112
info@edvantia.org • www.edvantia.org

Edvantia maintains policies and procedures that govern the Institutional Review Board in ensuring the welfare and protecting the rights of human subjects who choose to participate in research and evaluation activities.

This publication is based on work sponsored wholly or in part by the Institute of Education Sciences (IES), U.S. Department of Education, under contract number ED-01-CO-0016. Its contents do not necessarily reflect the positions or policies of IES, the Department, or any other agency of the U.S. government.

Online, Standards-Based, Formative Assessment Conference Proceedings

Mary Axelson

December 2005

Appalachia Educational Laboratory (AEL) at Edvantia

Contents

Introduction

> *You can't wait every year and hold your breath to know how the students are doing. The sooner you have that information, the better.*
>
> Mary Yakimowski, Director of Assessments, CCSSO

Online assessment of any type is attractive to educators because it allows large amounts of student data to be gathered and accessed quickly. Indeed, online, summative assessments—generally end-of-course, state-level, and placement exams—are making the annual student testing ritual a meaningful audit instead of what some call an "autopsy" because data reports arrive too late to have any impact on student learning. Yet educators now want meaningful data sooner and more regularly to plan instruction and intervention. That's where online, formative assessments come in.

In June of 2005, the Appalachia Educational Laboratory (AEL) at Edvantia, Inc., and the Council of Chief State School Officers (CCSSO) held a one-day symposium prior to CCSSO's National Conference on Large-Scale Assessment. The purpose of the event was to discuss issues related to standards-based, formative assessment delivered online. The Institute for the Advancement of Emerging Technologies in Education (IAETE) at AEL coordinated the symposium. The large number of attendees, as well as their comments throughout the day, clearly indicate that the topic resonated with states and districts; the results of CCSSO surveys of issues important to its membership also indicate great interest in the topic.

The event was the fifth in a series of annual assessment symposia sponsored by IAETE at AEL. In the fall of 2000, the Institute sponsored a symposium in Washington, DC, on the role of technology in large-scale assessment. In 2002, IAETE presented *Assessments that Empower Success: The Role of Technology* at the National School Boards Association's Technology + Learning Conference in Dallas, Texas. That event highlighted recent research in cognitive science and the implications for assessment that are identified in the National Research Council publications *How People Learn* and *Knowing What Students Know*. In 2003, IAETE offered a presession symposium at the annual meeting of the American Educational Research Association titled *Toward a National Research Agenda: Improving the Intelligence of Assessment through Technology*. In 2004, IAETE teamed with CCSSO for the first time to offer a full-day session, *Technology for Assessment: Tackling the Policy Issues,* at CCSSO's National Conference on Large-Scale Assessment.

Participants in the earlier symposia questioned the value of investing substantial time and money in annual summative assessments for accountability in light of the small impact they have on classroom instruction. Many feared that the broad but shallow content measured by these assessments would eliminate classroom explorations of any substantial depth. Responding to this sentiment in 2002, Chris Dede, Wirth Professor of Learning Technologies at the Harvard Graduate School of Education, advised

> We are in a "reform" movement where powerful methods of teaching/learning are harder to use, due to flawed standards and tests. The only way to improve this situation is to give people something to move toward—not something to move against—because then we'll just react away from what we have now into some other flawed method of reform.

Formative assessments offer something "to move toward." Using information and communications technologies, formative classroom assessments can reveal a timely picture of the subject matter students master, as well as the source of cognitive shortfalls, in time for teachers to plan and carry out interventions. Additionally, many promising approaches to assessment that use technology, including concept mapping,[1] simulations, and role playing, may well find more fertile ground for their initial growth in formative assessments.

The Day's Format

Three four-person panels addressed three issues key to online, formative assessment, with each panel taking on a different issue. Rather than make presentations, panelists shared issue-based illustrations from their own experiences. Panelists then took questions from conference participants. Table discussions and a reporting out from each table followed. The three topics and key questions explored by the panel were

1. **Infrastructure.** How can a state that can offer its current online assessments only once a year to a fraction of its students, because there aren't enough computers in the schools or enough bandwidth to support them, provide multiple testing events throughout the year in every class in every school?
2. **Human resources.** Technology infrastructure and expertise alone are incomplete solutions. Many people are needed to make the technology happen and to use the data that the technology can provide. How can states and districts secure and retain the technology, curriculum, and data specialists required to reap the benefits from a formative assessment system? What types of professional development should be offered to help educators develop both the technical and assessment proficiencies to capitalize on the promise of online formative assessment?
3. **Looking ahead.** How will the data be used once they are available—now and 10 years down the road? How can formative assessments be designed to accurately predict student needs and then be translated into effective classroom instruction? If formative assessments can become valid and reliable measures of student performance, how would that affect current assessment and accountability systems? Might current systems be replaced?

Summaries of the table discussions revealed participants' wide experience as well as their frustrations with the challenges of creating formative assessments as part of a wider assessment system. The symposium was a day full of highly informed questions—some were answered; some were left with the promise of further exploration.

Infrastructure Panelists

- Tony Alpert, Manager of Assessment Reporting, Office of Assessment and Information Services, Oregon Department of Education
- David Couch, Director, Division of Planning Services, Kentucky Department of Education
- Steve Henry, Director of Research, Evaluation, and Assessment, Topeka Public Schools, Topeka, Kansas
- Angie Shook, Executive Director of State Testing, Office of Accountability and Assessments, Oklahoma State Department of Education
- John Ross, Senior Research and Development Specialist, Edvantia (moderator)

Human Resources Panelists

- Sharron Hunt, Director of Testing, Georgia Department of Education
- Richard Schley, Educational Technology Specialist, Virginia Department of Education
- Jan Sheinker, Education Consultant
- Cindy Simmons, Director, Office of Student Assessment, Mississippi Department of Education
- Art Halbrook, Senior Associate, State Collaborative on Assessment and Student Standards (SCASS), CCSSO (moderator)

Looking Ahead Panelists

- Anita Givens, Senior Director for Instructional Materials and Educational Technology, Texas Education Agency
- John Poggio, Codirector, Center for Education Testing and Evaluation, University of Kansas
- Brenda Williams, Executive Director, Office of Technology and Information Systems, West Virginia Department of Education
- Phoebe Winter, Education Consultant
- John Ross, Edvantia (moderator)

The symposium also included a lunchtime gallery walk, which allowed participants to see technology-based, formative assessment products currently (or soon to be) on the market. Vendors were CTB McGraw Hill, Educational Testing Service (ETS), Learning.com, Pearson Educational Measurement, ThinkLink, and Wireless Generation.

A Common Vocabulary

> *At our table, and I'm sure within schools and states, we're using the same words to talk about different things . . . so, for example, formative could mean to some people diagnostic assessment. It could mean instructionally embedded assessment; it could mean mini state test. We need some definitions.*
>
> Phoebe Winter, Education Consultant

The day's discussion quickly revealed the need for a common vocabulary. Attendance was evenly split between assessment and technology professionals, all of whom already knew the difficulties of communication. Participants were, however, surprised to find that those with the same backgrounds from different states used terms differently.

Assessment types are largely defined by purpose. In her introductory remarks, Edvantia CEO Doris Redfield defined the different types of assessments. Many states offer *benchmark assessments* every six or nine weeks that purport to predict how students will perform on summative assessments. *Summative assessments* cover the entire scope and sequence of a course and are used to judge student mastery of the entire course content. Another example of summative assessment is the familiar, annual, high-stakes assessments administered by states and used to measure adequate yearly progress, as required in NCLB. Traditionally, these assessments have not helped teachers adjust instruction for individual students because it takes too long for data reports to be returned to schools. *Formative assessment* is what teachers do on a regular basis to monitor student progress and modify instruction.

Participants also expressed confusion over the various technologies mentioned: namely, computer-based assessment, online assessment, and Web-based assessments. *Computer-based assessment* is the most widely used term. If a student is using a computer to view items and respond to them, it is a computer-based assessment. The software managing the assessment could reside on the individual computer or on one connected over a network, local or otherwise. An *online assessment* accesses the software through a network, with current assessment solutions often being delivered from a local or distant server. *Web-based* assessments are online assessments that can be delivered over the World Wide Web or a local area network.

The Demand for Formative Assessment

> *Districts have an exceedingly strong interest in formative assessment content, so strong that if states don't move to provide it, districts are likely to go to pretty extraordinary efforts to come up with it themselves.*
>
> Steve Henry, Director of Research, Evaluation, and Assessment, Topeka Public Schools, Topeka, Kansas

> *We do have certain school districts marching on with formative assessments; it is crucial to them. They would consider it much more crucial than summative, by the way, by far.*
>
> David Couch, Director, Division of Planning Services, Kentucky Department of Education

The statements above are just two of many made during the symposium expressing the strong desire for formative assessment. The enthusiasm, however, is paired with caution. Steve Henry noted that districts actually may not be in the best position to create standards-based, formative assessments. Angie Shook, executive director of state testing at the Office of Accountability and Assessments at the Oklahoma State Department of Education, observed that districts buying

assessment items are often oversold on the correlation between those items and state standards. The state, prohibited from recommending a vendor, cannot advise districts on such purchases. David Couch expressed concern about equity issues created when wealthy districts pursue formative assessment independently.

Additional problems arise when, in the absence of formative assessments, educators seek formative information from summative assessments. Tony Alpert, manager of the Assessment Reporting Office at the Oregon Department of Education, explained that Oregon's assessment advisory committee comprised of district assessment staff proposed that strand data for individual students not be displayed.[2] Such groupings of questions on a particular topic have a higher standard error than composite scores and are not a reliable measure of a student's ability. Explains Alpert, "People are making curriculum decisions based on a single point in time and a test that covers a huge array of content." Even so, teachers are enticed by the prospect of formative assessment, and they are using the results of summative assessments, designed for making school-level decisions, to make decisions about individual students—right or wrong. Yet the state is reluctant to take that data away from teachers without replacing it with something more valid.

Across the board, the ultimate goals of the participants at this symposium were (1) to create comprehensive assessment systems that include both formative and summative assessments, (2) to foster an understanding among teachers of the unique roles of formative and summative assessments, and (3) to help teachers to use formative data to inform instruction. While formative and summative assessments were once seen as an either/or proposition, said panelist Jan Sheinker, an education consultant that has worked at the district, state, and national level, there is now a growing understanding of their distinct roles in a larger system. Sheinker anticipates systems in which states "build the online, formative assessment as they are rolling out the revisions to their content standards."

Infrastructure

> *At the elementary level, our district made a commitment a number of years ago to distribute computers in classrooms rather than have labs. And there was a philosophy behind that to promote real integration of the technology with instruction on an ongoing basis—but it does not work well for computer testing to have 4-5 computers in a classroom.*
>
> Steve Henry, Director of Research, Evaluation, and Assessment, Topeka Public Schools, Topeka, Kansas

> *We dedicated 75 percent of our bandwidth . . . to make sure the students taking the test got priority and everybody else got put in another lane . . . . During a two-week period in our state, that was not a big deal . . . you do that during the course of the whole year [and it's not] a big problem.*
>
> Tony Alpert, Manager of the Assessment Reporting Office of Assessment and Information Services, Oregon Department of Education

Defining and Funding the Technology Infrastructure

Online assessment of any kind requires that students have access to computers or other computing devices in the necessary time frame, that those devices have operating systems and browsers that support the assessment, and that there is adequate bandwidth to deliver the assessment. Ideally, this should all happen without disrupting computer access needed for instruction. As Couch said, "No matter which way you do that, formative or summative, you have to deal with almost the same technology issues from an infrastructure perspective."

States that have had success with online assessment have typically established the infrastructure first and, most likely, long before using it to deliver assessments. Said Richard Schley, an educational technology specialist with the Virginia Department of Education, "Before we could talk about online assessment, we had to build a powerful statewide infrastructure." The process, panelists made clear, often begins in the legislature, with funding that enables states and districts to build the infrastructure. Kentucky's infrastructure funding is divided across federal, state, and local sources. It takes, said Couch, "about $122 million annually to operate, maintain, and do incremental replacements at a very conservative rate." He expressed concern about the possibility of cuts to e-Rate funding, a move that could result in a loss of $20 million a year for his state.

Anita Givens, senior director for instructional materials and educational technology at the Texas Education Agency, speaking in the Looking Ahead panel, described an array of funding sources and the related grant requirements that had to be managed for a Texas pilot on technology immersion. Said moderator John Ross of Edvantia of the dizzying mix, "That's a funding model that I think a lot of schools don't understand—that is, pulling money from different pots together to generate a larger system."

In addition to funding, the location of computers in a school building is a serious consideration; indeed, it can be as significant as the technical capacity of technology staff. Elementary schools tend to have computers placed in classrooms, while many high schools favor the centralized placement of equipment in computer labs. Middle schools vary.

It was the general impression among panelists that it is easier to administer high-stakes summative assessments in labs because it is easier to monitor students and ensure a more secure physical environment. For this reason, many online testing pilots began with high schools. In response to the need for delivering summative assessments online, many elementary schools

simply gather computers together from multiple classrooms to make a temporary lab. That solution is not as practical for formative assessments, however, because formative assessments are administered frequently throughout the school year, sometimes once a week or more. Computers have great instructional value in the classroom, and panelists were not eager to discontinue efforts to place them there. Constantly moving equipment for frequent assessments is not practical.

Even if the computers are permanently placed in one location, they may not be available for formative assessment. In Kansas, for example, the labs and bandwidth are dedicated to course delivery. Schools there have difficulty scheduling summative assessments. Mobile labs have proven to be effective solutions to these complex problems in some states. Panelists also anticipate a role for handheld computing devices.

An online assessment needs to reside on a powerful and secure server—or servers. Panelists described states' broad array of approaches to this need. Kentucky's district technology staffs are overwhelmed, said Couch. Though they once would have fought state control over their work (some districts may view the state hosting an assessment as outsourcing to the state), they are now happy to have the state host assessment.

In the past three years, the Oregon Department of Education has had a budget cut of 30 to 40 percent. Alpert said the department is looking to its education service agencies and the districts themselves to host those services. Vendors currently host their assessments. The Georgia Online Assessment System is also hosted by a vendor.

The Receiving End: Technical Surveys

State technology leaders need some understanding of capacity at the district and school levels. Collecting technical readiness surveys from each school in the state was a common practice of all panelists. When Kentucky, for example, was preparing for an online, large-scale assessment, the state department of education conducted an online readiness survey and an inventory of workstations. The agency learned that school connectivity is half fiber and half T1, with T1 lines connecting the districts to the state. The picture of a robust technology infrastructure agency originally envisioned by staff faded a bit more with the realization that 75 percent of the workstations on the other end of those network connections are 6 to 12 years old.

Kentucky has statewide standards for technology products, so the advanced age of school computers was not a factor when the state considered the capacity of browsers, processor speeds, or operating systems. However, when Oklahoma went completely online with a seventh-grade geography assessment in the 2004-2005 school year, its survey of schools revealed numerous compatibility problems with various Internet browsers and operating systems. The vendor worked with schools to update those components. Oklahoma's survey did confirm the good news that bandwidth was sufficient. About 7 percent of schools had difficulty testing more than 20 students at one time, so students in those schools took tests at their local technology center.

Besides securing compatible software, hardware, and networks, it is also advisable to ensure that students are ready for the online medium. It is essential that students regularly use the required technologies of assessment for learning so that lack of technical proficiency does not hamper their ability to demonstrate their understanding of content. In some states that have online summative assessments, such as Virginia, teachers are using for instruction new interfaces and tools that were developed for assessment, to help students become better acquainted with the technology.

Item Banks

Valuable formative assessment requires items of the same caliber as those used in the summative assessment. Sheinker criticizes some formative assessments, explaining, "In their formative assessments, the student never gets anything resembling the level of content and cognitive challenge that the items on the large-scale assessment will expect. So that is really a problem." Correcting the situation requires items that have been aligned to state standards, field-tested, and reviewed for reliability, validity, and bias.

Formative assessment, as envisioned in this symposium, is a formidable beast that needs to be fed. Released state items, purchased items, or items developed by teachers are all possible sources to supply the necessary item banks. States may also benefit from a structure that allows them to share items. Virginia's item bank allows teachers to contribute formative items that, if accepted, are then sent through a review process before being made available across the state. Districts can also enter items for their own benchmark tests.

In Mississippi, said Cindy Simmons, Director of the Office of Student Assessment at the Mississippi Department of Education, "Our teachers can create items now. They can use the items in the system or create their own, but there is no oversight. We are in the process now of creating a state-level user hierarchy so that any teacher-submitted items can be reviewed by people at the state level and added to the item bank that exists." Mississippi and other states have also discovered that the alignment of purchased items to state standards must be carefully reviewed.

Many states, such as Kansas, are now evaluating how to package and distribute formative assessment items. Steve Henry (Topeka Public Schools) discussed Topeka's effort to provide teacher workshops focused on item development. Topeka's computer-based system stores items and repackages them in what Henry called a *testlet*. These mirror the state reading assessment, with three comprehension items, one item on literary structure, and one vocabulary item. At this point, teachers download the items and make overhead transparencies to present the items to students.

Georgia, explained Sharron Hunt (Georgia Department of Education), has a more high-tech model. The Georgia Item Bank, with approximately 12,000 items, is the source of both the state's formative and summative online assessments. They are made available to different user groups based on four levels of access:

- Level 1 can be accessed by students, parents, and the public, at school or home. The items are organized as prebuilt assessments around the domains within a subject area. It is currently operational.
- Level 2 is reserved for teachers, who can access the entire system and create their own formative assessments. This build-your-own option is operational. A prebuilt option is proposed for the 2006-2007 school year.
- Level 3 is a secure benchmark level designed to predict performance on summative assessments. It became operational in the 2006-2007 school year.
- Level 4 will generate the state summative assessment across all grade levels. It is three to five years away from being operational.

Security

Summative assessments can bear significant consequences. They can determine whether students move from grade to grade or even graduate as well as determine whether schools and districts meet performance standards, including adequate yearly progress (AYP). These high stakes generate many concerns about test security. The ability to access items used for both formative and summative assessments from the same bank adds to the worries. An error while switching vendors in Georgia did lead to the release of a summative assessment item, so the fears are not easily dismissed.

Hacking into the testing system is but one concern. An additional concern is the possibility of students cheating or discussing items. Though the issue looms larger for summative assessments than for formative, it can't be dismissed for formative use. Formative assessments can employ the same security measures as summative assessments, such as changing small characteristics in an item that do not affect the construct.

Human Resources

> *What we discovered is, there is information in the large-scale assessment, but there is also information that cannot be measured there. So how do you help people to understand the relationship of the pieces of information within the assessment system? That is very core to the professional development in this process.*
>
> Jan Sheinker, Education Consultant

Introducing the panel on Human Resources, moderator Art Halbrook, senior associate of the State Collaborative on Assessment and Student Standards (SCASS), CCSSO, raised two major questions:

1. "How can states and districts secure and retain the qualified technology, curriculum, instruction, assessment, and data specialists required to reap the benefits from a formative assessment system?"

2. "What types of professional development can and should be offered to help educators develop both technical and professional efficiencies for capitalizing on the promise of online formative assessment?"

Both questions found their way into the conversation before the Human Resource panel made their appearance. It is an urgency caused, in part, by budget issues. Said Couch, in reference to Kentucky,

> Our average district has eight schools, and the average district in our state has 2 to 3 technology staff people, total. A company equal in size to a school district and with an equal number of computers would have 10 to 15 people.

The Seemingly Impossible Design Dream Team

Until recently, psychometricians, technologists, and content experts have most often settled for a "you-do-your-part, tag team" approach, rather than participating in a fluent exchange of thoughts. However, understanding and realizing the new possibilities available from online assessment demands a more unified team. Anita Givens, senior director for instructional materials and educational technology at the Texas Education Agency, cited "better partnerships with all of the players that have to come together to make these systems work" as her greatest desire for the future. She added,

> If we are going to change education practice, if we are going to change assessment practice, if we are going to change instructional practice, we have to get all of the players who are involved in those practices talking to each other and planning together.

This model of collaboration is one that some states are beginning to adopt. West Virginia's assessment design team, explained Brenda Williams, executive director of the Office of Technology and Information Systems at the West Virginia Department of Education, had what she called a *technology interpreter* to ease communications and identify technology issues for the assessment staff. She noted that it is common to have conversations in which nobody realizes that the terms they are using have different definitions for everyone involved. Givens credited a willingness to learn the jargon and the priorities of different domains for the success of a technology-based assessment project in Texas. Interestingly, simultaneous work at IAETE provides anecdotal testimony that designing an online assessment of eighth-grade technological literacy can open communications between assessment and technology departments. For more information on stories about technology and assessment teams working together to address online assessment of technology proficiency, refer to the article *Online Assessments of Technology Literacy: The Perfect Petri Dish* by Mary Axelson in IAETE's free, online publication *InSight* at www.iaete.org/insight/.

Virginia, explained Richard Schley, educational technology specialist with the Virginia Department of Education, managed to find a single individual with both a strong assessment background and a strong technology background. The two abilities, however, do not often

occupy a single mind and resumé. In Virginia, this person served as the project manager who oversaw the development of the Commonwealth's online assessment system. On the contrary, using a project manager who was outside both the assessment and technology departments worked well for Kentucky.

The more prevalent situation might be what Gretchen Ridgeway, a symposium participant from the Department of Defense Education Activity, observed at a field test site for the department's worldwide online assessment. She observed that the technical staff were more likely to administer the assessment than were designated assessment staff. Asked why, Ridgeway noted that the assessment staff were not comfortable with the technology. She said they were doing a good job, but in a separate conversation, Couch (Kentucky Department of Education) advised, "In no situation should your technology folks be leading the effort. You need a formalized governing structure."

The array of expertise required to create online assessment systems often requires states to outsource the project to vendors. Even then, warned the experienced panelists, states need to create an exceptionally clear RFP. Said Cindy Simmons, Director of the Office of Student Assessment at the Mississippi Department of Education, "When we issued the RFP for our student progress monitoring system, it really was a concerted effort. Our IT department that oversees all state agencies, plus our IS department within our education department, and the assessment department within the education department worked together on that RFP." Issuing the RFP is followed by a continued need for good communications among interested parties through the design, development, and deployment of assessment systems.

The need for good communication becomes ever stronger when states are working with multiple vendors on a single project—especially when that project occurs across multiple years. Mississippi experienced problems when software from different vendors could not exchange data. Sharron Hunt (Georgia Department of Education) built on Georgia's experience and recommended developing a transition plan for when one vendor leaves and another takes over. As the gallery walk of vendors at the event demonstrated, wonderful products are available and don't need to be reinvented. Customizing for a state, however, demands continued teamwork and communication.

Professional Development: Helping Educators Understand the Data

At this event and the preceding symposia it was observed that most teachers have not been expected to understand assessment. An area left to specialists, the subject is rarely even a required course in teacher training programs.[3] John Ross (IAETE moderator) candidly observed, "As a teacher, trained many, many years ago, when I got those scores, my job was to give them to the kids so they could take them home to their parents. And that's about all that I was trained to do."

But that situation is changing to an expectation that educators be "assessment literate." Teachers need assistance in creating items, and, more significantly, must be able to adjust instruction based on the data that returns from assessments. Additionally, teachers must be able to

effectively integrate the same technologies used for assessment into their instruction. As CCSSO's Halbrook pointed out, "Success is built on comfort."

A "train-the-trainer model" is an often-used system for tackling professional development. It is not, however, without its problems. Staff selected to attend the training sessions are not always the most qualified or prepared to take it back to their home school or district. Mississippi benefited greatly when the vendor took the most enthusiastic teachers to its headquarters for free, intensive training. Said Cindy Simmons, "The one thing we heard consistently [from teachers] was that they really didn't want trainers from the vendor." Simmons further said that the teachers wanted to be trained by familiar people who had similar experiences and understood the challenges they faced in their own schools. This type of train-the-trainer model seemed more relevant to the teachers, who were responsible for carrying out new policies and procedures and using the new, technology-based systems.

These training efforts tend to focus on a state's particular assessment software. There is not yet a major effort under way to teach the ins and outs of understanding assessment data. At the most basic level, states and districts see a need to ensure that teachers do not misuse data, as mentioned earlier regarding Oregon teachers' use of strand data. But teachers also need to be able to look at data or reports from formative tools and understand what kind of intervention or remediation is needed—at both the classroom and individual levels. Though panelists regarded professional development, preferably embedded in the classroom, as the obvious need, John Poggio, codirector of the Center for Education Testing and Evaluation at the University of Kansas, made a plea for improved reports:

> Somebody said, "Our standard error was 3.0, and when we did it on a subtest it, was 8.0." Now, I fully understood what was being said, but can you imagine reading that in a report and saying, "Make a decision about this child?" Somehow we have to do a better job . . . . Here's the question to ask yourself: Three years down the road, when the blush is off the rose, when the speed of the return of the results is second nature to everyone, what is going to sustain this? And I think it starts to relate to things like quality of reports. These things have to be understandable.

Georgia, said Sharron Hunt, has two kinds of training for computer-based assessments. One is technical training, which includes issues such as how to monitor progress or how to import student demographic data. The other is how to make assessment integral to instruction. Both are ongoing through regional education service agencies, using a train-the-trainer model. Says Hunt,

> We are in the second year of [the training], so I would hesitate to say how successful it has been. But, if usage on the system might be a measure, in March and April of '05, we had four million test events recorded. About two-thirds of those were from the teacher level of the bank and one-third of those were from the student level of the bank. Because our student/parent level is increasing, our third model of training is one in which we have partnered with the state PTA.

Student Preferences

Another human resources issue to consider relates to students. Several panelists noted that students appear to prefer online testing. Oregon students demonstrated a preference for computer-based testing. In Oregon, teachers and administrators have additional motivation to take the online option because this option's administrative ease has created a policy of giving students three chances to pass, thus significantly helping schools demonstrate adequate yearly progress.

Kentucky also surveyed students after they completed a large-scale online test. Responses showed that

- 80 percent preferred taking a test online to taking it on paper
- 80 percent said they were able to focus more easily using a computer than using paper
- 83 percent thought the online test was easier to take
- 87 percent said they want to take the test online next year

David Couch (Kentucky Department of Education) said the improved ability to focus on the test was evident from simple observation. Those on a computer looked at the screen; those with paper and pencil looked around the room.

Looking Ahead

> *Technology offers an opportunity for us to do something different with formative assessment. I really believe that. And that has to do with using more performance assessment, using more constructed response assessment, assessing more frequently.*
>
> Jan Sheinker, Education Consultant

> *A danger of any of these things from an equity standpoint is that we end up doing a really nice job for what I used to like to call "the psychometrically perfect child," but the rest of the students are left out.*
>
> Phoebe Winter, Education Consultant

Anticipating the possibilities with formative assessment in the future, panelists and participants look forward to new kinds of questions, new technologies for delivering items and processing student responses, benefits for students with disabilities, and pairing instructional strategies with assessment performance.

Comparability and Question Types

Every IAETE/AEL symposium on online assessment has brought up the issue of how student performance on online assessments compares to performance on paper-and-pencil assessments. John Poggio responded to the inevitable request by saying, "I think we lose a lot of time dealing with comparability issues. Comparability is important only during the time of transition. Who made paper and pencil the gold standard?" Indeed, the question of comparability studies was raised by attendees during the question-and-answer sessions, with several in attendance noting organizations that were undergoing, or that planned to complete, comparability studies that could be shared on state or organization Web sites.

Oregon, however, has found that some items function differently by subgroups of students with the two forms of media. "We haven't cracked the code on why," said Tony Alpert. "We are not seeing an overall difference on test scores by mode of delivery, but the particular item functioning is a concern." Phoebe Winter has found it important to ask if students think differently with the different formats. "How students think might be different on a computer than [with] paper and pencil." There is no research-based answer, but the observation of increased focus with the computer does suggest a fundamental difference in interaction.

The group clearly held a common hope that technology can create superior items or performance tasks. The multiple-choice assessment item was widely viewed as "the lowest common denominator." But many of the panelists offered hope that new technologies would provide improved methods, tools, and processes for assessment. Said Phoebe Winter,

> When someone said the infrastructure wasn't ready yet, at first I thought, "Ah, that's too bad." But then I thought, "No, that might be a really good thing." Because that means we can take advantage of the research that's just happening and maybe even do some research. And this research could incorporate ideas of student learning, cognitive psychologies, [and] psychometrics into technology-based assessment.

And so the quest evolves from different item types to new ways to target the depth of student understanding. As Jim Pellegrino, distinguished professor of cognitive psychology and education at the University of Illinois at Chicago, explained at the 2002 symposium,[4] "We know from cognitive science that assessment has to move beyond assessing discrete bits and pieces of knowledge. We must move toward assessing more complex aspects of knowing and understanding."[5]

Oregon is working toward scoring its writing exam through artificial intelligence. At the time of the symposium, each writing sample had two human scorers, with a third person reconciling the differences. For the 2005-2006 school year, Oregon will score a portion of the online tests in the background with artificial intelligence and compare those scores to human rates. Despite some trepidation about the approach, the potential reduction in costs is driving the state forward.

Discussion of the opportunity to score new types of questions (such as essay and constructed response) and simulation or performance-based tasks circled back to issues related to

infrastructure and raised anew the question of bandwidth. Certainly, moving beyond multiple choice will require a bigger pipe in both directions.

John Poggio entered a plea for adaptive testing (testing that is tailored to the ability level of the test taker), saying, "We are discussing online testing and formative testing as though the world was built for the paper-and-pencil universe. It will be *adaptive* . . . . We should be thinking about *adaptive* formative tests and *adaptive* summative tests." He pointed out that the infrastructure demands can be reduced considerably when every child stops taking a 75-item, fixed-form test, which, ultimately, can be reduced to a 20-minute adaptive test. Adaptive testing also has the security benefit of offering a different series of questions on-screen in a lab. Currently, adaptive testing is not allowed by the U.S. Department of Education for assessments that are used to determine measures of adequate yearly progress.

New Devices

Frequent, formative assessment requires more than bandwidth; it requires frequent student access to a computing device. Many panelists agreed that a desktop might not always be the best choice. Different devices may be more appropriate depending on the grade level and assessment purpose.

Texas has run a highly successful pilot using handhelds to capture performance evaluations of early reading.[6] Said Anita Givens (Texas Education Agency), "We were suffering from success almost immediately. The teachers loved this new way of giving assessments." Prior to the pilot, teachers throughout the state administered mandatory reading fluency assessments, such as the Texas Primary Reading Inventory (TPRI), which presented oral activities for children with responses recorded by teachers on paper. When all assessments were completed, teachers were faced with stacks of paper that had to be summarized by hand or entered the data into the computer. Administrators were faced with similar stacks of paper from multiple campuses or classrooms. Although accuracy was never an issue with this method, workload was, and so was timeliness and relevancy of the data. Said Givens, "Sometimes they got that data back and they looked at it and they did things with it, but, for the most part . . . the vast majority just turned it in because they had to. They didn't really use that data to guide instruction."

The handheld pilot addressed two goals of the Texas program: (1) to tie instructional strategies to results, and (2) to reduce the time required for the overall assessment. With the handheld, the assessment experience stays the same for the child, but teachers enter results live, as the child is responding. The handheld is then synched to the computer and uploaded to a secure database. Teachers receive results almost instantaneously. They can then consult information provided by the state that links areas of student deficiencies to instructional strategies recommended to address them. Additionally, the handheld can prompt teachers with guidelines for administering the assessment, and it eliminated the need for a separate timer.

Creating the pilot, explained Givens, required finding a vendor and a group of teachers willing to try the new technology-based method for administering the familiar assessment. They started with three schools and about 25 teachers and gradually expanded those numbers as they

improved the software. As that happened, they realized how important was that everyone involved with the reading program receive training. “Professional development was the key,” Given said. “We had to do the professional development for *every* person involved in the project: first on the technology, then on the software and assessment tool, then on putting those two together in the classroom.”

In the second year, the state expanded the program to 100 elementary campuses and offered an early-adopter program for the many other schools that wanted the handheld option enough to pay for it themselves.

By the end of the second year, teachers were saving an average of four and a half hours of data entry time per administration. Given that a teacher might administer the TPRI up to three times a year, the handhelds freed up an extra 15 hours per teacher that could instead be spent reviewing the data and designing appropriate instructional strategies. The state did not need to push this project to schools because teachers were asking for it. Early on in the project, teachers also asked for a Spanish version. That version, Tejas Lee, is now also available to schools. What started as a pilot project is now in use in schools across the nation. In a new project, Texas is now piloting an early-grade math assessment using the same platform and handheld device.

Tony Alpert (Oregon Department of Education) expects handhelds to resolve some of the bandwidth issues presented by formative assessments competing with instructional use of computers. He notes that Oregon already uses the devices for observations, so it seems reasonable to use them for testing. “We do see a tremendous growth in that industry, not just from the teachers but from principals and students as well,” he said. Kentucky is also enthusiastic about the price point and functionality of handhelds.

Beyond Handhelds

The ability to capture student responses digitally is central to the ability to analyze data. Panelists described the current popularity of wireless polling devices that allow teachers to collect instant results to questions and adapt the day’s instruction accordingly. These devices range from a special device resembling a TV remote control to PDAs. Some whiteboard systems also support this functionality.

Tablet computers are another consideration for new devices to support assessment. Their mobility, observed Poggio, makes them an ideal tool for capturing performance-based data in science and math. Their capacity for handwriting recognition may also hold benefits for assessment.

Pairing Assessment Results with Instructional Strategies

The Texas Reading Inventory, in which an analysis of student performance automatically links teachers to instructional strategies, is one example of efforts to pair instructional strategies with assessment results. In West Virginia, too, said Brenda Williams, the state recognizes a tie to instruction as vital to the cycle. The state is creating a "curriculum matrix" to tie deficits to curriculum pieces such as online simulations and higher order thinking skills resources. "We're going to have that match, so that teachers not only have the resources to do the assessment but also the staff development to understand what that means. They will also have the other pieces and resources to help students get the piece that they're missing."

Ideally, assessments can identify not only gaps in student performance of the standards but the best way to address them. Phoebe Winter (consultant) identified the hope that formative assessments could identify individual learning styles so that not only content but also format is appropriate for each student.

Equity and Universal Design

In Kentucky, assessing students with special needs originally drove the creation of an online summative assessment. The 2005-2006 school year is the first time online assessments have been opened up to other students. Many students with disabilities use technology to interact with the world, and it is an obvious tool for helping them receive and respond to assessment items.

Separate work at the Appalachia Educational Laboratory at Edvantia underscores the importance of designing assessment for students with disabilities from the ground up.[7] Oregon is one of several states to discover the difficulty of retrofitting an online assessment for disability access. Alpert said a panel is developing modifications for students who need them, and most require more bandwidth. Having the computer read questions aloud, for example, is a frequent need.

Winter hopes not only for accessible assessments but for assessments that identify disabilities and abilities for all students. She thinks that technology can help create formative assessment systems that help teachers understand different learning modes and response modes, as well as how different students learn.

Refining the Cycle

"Maybe someday," said Winter, "we can actually incorporate information we get in the classroom about students into our large-scale systems for accountability and evaluations." The ultimate goal is not to provide a variety of assessment but, as Winter's said, to "develop the research base so that you are connecting instruction and assessment in a way that will really further instruction and learning." A participant from one of the table discussions remarked that the group expected embedded assessment to emerge as the norm for both summative and formative needs.

That vision remains futuristic. As Anita Givens pointed out, in another pilot, the Texas Department of Education had to be more flexible with a requirement for every teacher in every grade to use specific formative assessment tools because teachers thought they would take time away from preparing for the state's summative tests. She, too, anticipates that a "seamless system" can one day replace competing demands.

Where Now?

> *Communication is not only critical within the state and within our departments but also across the states. Some of you are starting other projects, and you may have other pieces we could fit together. It's time critical for what's happening in the classroom.*
>
> Brenda Williams, Executive Director, Office of Technology and Information Systems, West Virginia Department of Education

The symposia series has helped the Appalachia Educational Laboratory at Edvantia sculpt a scope of work. Previous symposia, for example, identified the importance of formative assessment and the need to broaden understandings of complete assessment systems. But the importance of this work is lost unless a dialog is created, not only across departments within a state but across states.

A frequently asked question during the symposium was best expressed in the table reporting sessions: "How to avoid re-creating these systems 50 times? We need a think tank," read the notation. If not a think tank, then at least a database or clearinghouse of state practices seems to be an essential starting point. As Williams explained, West Virginia learned much from Texas when it started a handheld reading inventory, but the program could not be duplicated exactly because the state had different needs and opportunities. Still, much was learned by sharing, and a customized approach was developed.

Less apparent is the opportunity to tie emerging research to the creation of different pieces of the assessment cycle. John Ross, director of IAETE and symposium moderator, accepted the challenge to share stories of experiences and new findings and placed some of the burden squarely on the shoulders of the organizations represented at the symposium. He noted that symposium proceedings would be posted on Edvantia's regional educational laboratory Web site and that answers to questions raised during the day could be addressed not only by the lab but by every organization present. States and vendors completing comparability studies were challenged to make those studies available and to share their findings through venues like the symposium. The day was successful from the their standpoint that so many state educators and vendors spent a good deal of time talking and sharing, voicing their concerns, and raising difficult issues. The issues raised were ones better addressed by all of the groups, rather than by any one group trying to offer all the answers.

John Poggio (University of Kansas) made a summary observation about the increased importance of the roles assessment and technology staff have come to play in terms of school reform:

> For the moment, let's take a step back and recognize that what we are all about is assessment-driven reform. Make no mistake about it. Ten years ago, in the job you are presently in, you were just a nuisance. Today you are the focal point of change driving all institutional reforms in your states.

---

[1] CRESST

2 Within Oregon's tests are *strands*, or sections of questions for a specific subject area. The standard error on a strand (plus or minus 8 scale points) is higher than that of the composite score (plus or minus 3).

3 M. A. Axelson. *Technology for Assessment: Tackling the Policy Issues* (Charleston, WV: AEL, 2005), page 10.

4 M. A. Axelson, T. M. McGraw, and S. McEntee, eds, *Assessments that Empower Success: The Role of Technology* (Charleston, WV : AEL, 2002).

5 Assessment that Empower Success: The Role of Technology, Conference Proceedings, http://www.edvantia.org/publications/pdf/Conference%20Proceedings%201.pdf, page 9

6 Wireless Generation created mCLASS™:TPRI® in partnership with the University of Texas Health Science Center at Houston, the University of Houston, the Center for Academic and Reading Skills (CARS), and the Texas Education Agency to save teachers time in conducting the assessment and to make it easier to collect and analyze valuable student performance data.

7 Axelson, Mary. (in press). *Maximizing the effectiveness of online accountability assessments for students with disabilities.* Edvantia: Charleston, WV.